Remarks

This is filed in response to the first Office Action in the above-cited case, which rejects the pending claims as allegedly anticipated or obvious in view of Lipkin (US 2002/0049788A1), Schuetze (US 2003/0074369a1) and/or Hsiung (US 2003/0109951A1). The claims are amended as indicated above and, as discussed below, are patentably from that cited reference.

Specification Amendments

The specification is amended, at page 10, to accommodate an error in the labeling of Figure 4. That error resulted in subject 406 being labeled "408" and object 408 being labeled "406." At page of the specification is further amended to correct the erroneous reference to element 420 as a "subject" rather than as an "object." No new matter is added.

Claims 1 – 8 and 14 – 16 are Patentably Distinct from the Art

Claims 1 – 8 are rejected as anticipated by Lipkin and/or obvious over Lipkin in view of Schuetze. For the reasons below, that rejection should be reconsidered and withdrawn.

Claim 1 is directed to digital data processing methods for enterprise business viability. The method includes transforming data from a plurality of databases into resource description framework (RDF) triples, storing those triples in a data store, and traversing one or more of the triples in the data store using self-adapting, genetic algorithms in order to identify data responsive to a query.

Claim 1 is amended to include the features of claims 4 – 6, specifically, reciting that traversing the database includes performing a plurality of searches on the data store, each search utilizing a different methodology, comparing results of plural such searches, and discerning from the comparison one or more of the searches that produce better results and re-performing those one or more searches on the data store with any of additional terms or further granularity.

Claim 14, which has been amended to include the limitations of the claim from which it previously depended, parallels claim 1, reciting that the RDF triples are stored in a data store.

Lipkin is directed to a web content platform with a repository of metadata in RDF form imported from external sources. *See*, Lipkin at column 62, ¶ 938. That publication teaches that

the repository is searched by applying an SQL-like search. *Id.* at column 70, ¶¶ 1095 1100. Nowhere does Lipkin teach or suggest applying self-adapting, genetic algorithms against the repository in order to identify data responsive to a query. The Examiner admits this in ¶ 4 of the Office Action.

Contrary to the Examiner's assertion, Schuetze does not remedy the deficiencies of Lipkin. That secondary reference is directed to methods of identifying similarities among objects in a collection. The first passage of apparent relevance relied on by the Examiner, to wit, page 2, ¶¶ 24 – 25 (column 2, lines 47–67), teaches that image retrieval tasks can be handled by applying a weighted search to different image features, such as histogram, texture and shape. It odes not suggest self-adaptation but, to the contrary, that if the user changes the weights, he or she must run a new search.

> [0024] The text surrounding or associated with an image often provides an indication of its context. The method proposed herein permits the use of multi-modal information, such as text and image features, for performing browsing and retrieval (of images, in the exemplary case described herein) This method is applicable more generally to other applications in which the elements (e.g., documents, phrases, or images) of a collection can be described by multiple characteristics, or features
>
> [0025] One difficulty in the use of multiple features in search and browsing is the combination of the information from the different features. This is commonly handled in image retrieval tasks by having weights associated with each feature (usually image features such as color histogram, texture, and shape) that can be set by the user. With each revision of the weights, a new search must be performed. However, in employing a heterogeneous set of multi-modal features, it is often difficult to assign weights to the importance of different features. In systems that employ metadata, the metadata usually has finite, discrete values, and a Bool-

The second passage of apparent relevance relied on by the Examiner, page 3, ¶ 31 (column 2, lines 8 17), discusses an apparent alternate approach in which like data is clustered, presumably before it is searched. In the event that the Examiner had intended to rely on page 3, ¶ 26 (column 1, lines 8 – 17), that passage discusses repeating a given methodology over and over to enable the user to find an image of interest:

> [0026] Current image retrieval systems (such as QBIC, Virage, and Smith & Chang) commonly display a random selection of images or allow an initial text query (such as a starting point. In the latter case, a set of images with that associated text is returned. The user selects the image most similar to what they are looking for, a search using the selected image as the query is performed and the most similar images are displayed. This process is repeated as the user finds images closer to what is desired. In some systems, the user can directly specify image features such as color distribution and can also specify weights on different features, such as color histograms, texture, and shape. In web pages, text such as URLs may also provide clues to the content of the image. Current image retrieval technology also allows the use of URL, alt tags, and hyperlink text to index images on the web. One approach also attempts to determine for each word surrounding an image caption whether it is likely to be a caption word and then matches caption words to "visual foci" or regions of images (such as the foreground). The Webseek image search engine and MARS 2 allow for relevance feedback on images by marking them as positive or negative exemplars.
>
> browsing, searching, retrieving and recommending content in a collection of documents.
>
> [0031] Clustering of large data sets is important for exploratory data analysis, visualization, statistical generalization, and recommendation systems. Most clustering algorithms rely on a similarity measure between objects. This proposal sets forth a data representation model and an associated similarity measure for multi-modal data. This approach is relevant to data sets where each object has several disparate types of information associated with it, which are called modalities. Examples of such data sets include the pages of a World Wide Web site (modalities here could be text, inlinks, outlinks, image characteristics, text genre, etc.).
>
> [0032] A primary feature of the present invention resides in its novel data representation model. Each modality within each document is described herein by an n-dimensional vector, thereby facilitating quantitative analysis of the relationships among the documents in the collection.
>
> [0033] In one application of the invention, a method is described for serially using document features in different spaces (i.e., different modalities) to browse and retrieve information. One embodiment of the method uses image and

Neither of these passages (like that discussed above) go any distance toward teaching that self-adapting, genetic algorithms can be applied against a data store — much less, a store of RDF triples — in order to meet a query. More particularly, they do not teach that such algorithms can perform a plurality of searches on such a data store, where each search utilizes a different methodology. Nor do they teach comparing results of plural such searches, and discerning from the comparison one or more of the searches that produce better results and re-performing those one or more searches on the data store with any of additional terms or further granularity

Contrary to the Examiner's further assertion, nowhere does Schuetze teach or suggest comparing the results of multiple searches, discerning from that comparison which search produced better results, nor re-performing those searches on a data store with additional terms or further granularity. To the contrary, the passage at page 19, column 1, lines 16 -- 42 teaches that two groups of data (objects) can be compared by running the same search on both: not claimed

[0221] Another data mining application is the discovery of unusual objects. For example, in the discovery phase of a lawsuit, a law firm may only be interested in outlier documents, not in the large groups of similar documents that mostly contain boilerplate. Multi-modal clustering would identify the large groups of similar documents (e.g., because of shared boilerplate). Interesting document would then be among those that are most distant from the centroids of large clusters.

[0222] A data mining technique according to the invention compares two groups of objects by doing a multi-modal clustering for the first and then assigning the second group to the clusters of the first. This analysis technique has been successfully used to compare Xerox-base and non-Xerox-based users of the Web site and found surprisingly few differences mainly because Xerox employees are users of Xerox products and that is one of the main reasons to go to the external Xerox web site (to download drivers, look up product information, etc). One difference was that a higher proportion of Xerox users visited only one page, the Xerox home page. The reason is probably that many browsers of Xerox employees have the Xerox home page as their default page, so that the user automatically goes to the Xerox home page when starting up their browser and then moves on to a page on a different site. This example demonstrates the utility of multi-modal clustering for comparing different user groups.

The other passage relied on by the Examiner in this regard, on page 17, column 2, lines 31 – 40 teaches a formula for determining a cluster:

$$P(p|c) = \sum_{u \in c} \frac{1}{|c|} P(p|u)$$

[0209] where |c| is the total number of users in the cluster c. This visualization can be thought of as a "density plot." Intuitively, it answers the question of where a typical user from this cluster is most likely to be. In a presently preferred embodiment of the invention, all non-zero probabilities are mapped onto a scale from 0.3 to 1.0 so that even pages that are only accessed a few times by users in the cluster are clearly visible.

And, while the passage cited by the Examiner on page 3, column 2. lines 28 – 40 suggests an iterative approach to searching, it teaches that the single, same methodology of clustering is applied each time:

> The method takes advantage of multiple ways in which a user can specify items of interest. For example, in images, features from the text and image modalities can be used to describe the images. The method is similar to the method set forth in U.S. Pat. No. 5,442,778 and in D. Cutting, D. R. Karger, J. O. Pedersen, and J. W. Tukey, "Scatter/Gather: A cluster-based approach to browsing large document collections," Proc. 15th Ann. Int'l SIGIR'92, 1992 ("Scatter/Gather") in that selection of clusters, followed by reclustering of the selected clusters is performed iteratively. It extends the Scatter/Gather paradigm in at least two respects: each clustering may be performed on a different feature (e.g., surrounding text, image URL, image color histogram, genre of the surrounding text); and a "map" function identifies the most similar clusters with respect to a specified feature. The latter function permits identification of addi-

For these reasons, among others, the combined teachings of Lipkin and Schuetze fail to teach or suggest the subject matter of amended claims 1 and 14, which require *inter alia* traversing one or more of the triples in the data store using self-adapting, genetic algorithms by performing a plurality of searches on the data store, each search utilizing a different methodology, comparing results of plural such searches, and discerning from the comparison one or more of the searches that produce better results and re-performing those one or more searches on the data store with any of additional terms or further granularity. Accordingly, neither those claims nor the claims which depend therefrom are anticipated or rendered obvious.

Claims 22 – 25 are Patentably Distinct from the Art

Claims 22 – 25 are rejected as anticipated by Lipkin and/or obvious over Lipkin in view of Hsiung. For the reasons below, that rejection should be reconsidered and withdrawn.

Claim 22 is directed to a digital data processing method that includes transforming data from a plurality of databases into resource description framework (RDF) triples, where at least two of the databases are of disparate variety, and storing those triples in a data store. Expiry data is stored with at least selected ones of the triples. The method is amended to recite the steps of responding to a search performed against the data store, if data requested is not stored in the data store, by applying a query to one or more of the plurality of databases to obtain the requested data. The amended claim further recites responding to such a search, if the data is stored in the data store

and is associated with expiry information indicating that the requested data may have expired, by returning the requested data with a reduced confidence factor.

The Cited Art

The principal reference, Lipkin, is directed to a web content platform with a repository of metadata in RDF form imported from external sources. *See*, Lipkin at column 62, ¶ 938. Contrary to the pending claims, nowhere does Lipkin teach or suggest storing expiry data with at least selected triples in the data store. Nor does Lipkin suggest responding to a search performed against the data store, if data requested is not stored in the data store, by applying a query to one or more of the plurality of databases to obtain the requested data; nor, responding to such a search, if the data is stored in the data store and is associated with expiry information indicating that the requested data may have expired, by returning the requested data with a reduced confidence factor.

Contrary to ¶ 5 of the Office Action, Hsiung does not remedy this deficiency. The passages relied on by the Examiner in this regard have little to do with storing expiry data with data in a data store, much less, that contained in a store of RDF triples. Thus, on page 26, at column 1, lines 44 67, Hsiung merely provides:

> [0365] Monitors shall have the ability to specify the time from when the sensor should begin monitoring. The user shall indicate if the change is for this monitoring session only, or whether this change should be remembered for future When a sensor is "opened" (e.g., displayed in a system view), the data is displayed from a point in time based on either this user's defined preference, or the "Default Time Frame" stored with the sensor if the user has not defined a preference. The user should select the starting time relative to the current time, and specify the time in hours and minutes. If the time period chosen is earlier than the time the current sensor was opened, the sensor data must be calculated from the starting point to the present, using the set refresh rate (the rate at which the model normally updates).
>
> [0366] The Print View allows monitors to use their browser's print function to print the system view. No special requirements.
>
> [0367] Monitors shall have the ability to save the system view (i.e., the graphical view) in a file. The system view should be saved in a standard graphic form for easy input into a MS Office document (e.g., Word, PowerPoint, etc.).

While, in column 2, lines 1 – 14 of that page, Hsiung provides:

> [0368] When a system view containing a model is opened, all models in the view are displayed with the configuration saved with that system view or with the model's defaults. The Software enables examination of individual analyses comprising the model.
>
> [0369] Monitors shall be able to click on any individual point in a model to get additional detail. When an individual point is selected, an appropriate graph/plot for that point should be opened in a separate browser window. The next chart to be displayed will vary depending on the analysis being viewed, and the level of the chart being viewed. In order to change current & default time frame, the requirements are the same for models as they are for sensors.

Neither of these passages can be construed as storing expiry data in the manner recited in the pending claims — nor is it at all evident how those of ordinary skill in the art might combine Hsiung's teachings with those of Lipkin to discern such a capability.

For these reasons, neither Lipkin, Hsiung, nor the combined teachings of both references, teaches or suggests the subject matter of pending claims 9 – 13 and 17 – 25. It is therefore requested that the rejection of those claims be withdrawn.

Conclusion

This responds in full to the pending Office Action in the above-cited case. The specification is amended to correct minor clerical errors and the claims are amended to clarify patentable distinctions over the cited art, Lipkin, Schuetze and/or Hsiung. In view hereof, the Applicants request that the rejection be withdrawn and that application be passed forward to issuance.

Respectfully submitted,
NUTTER, McCLENNEN & FISH, LLP

David J. Powsner
Reg. No. 31,868

Attorney for Applicant
World Trade Center West
155 Seaport Boulevard
Boston, MA 02210-2604